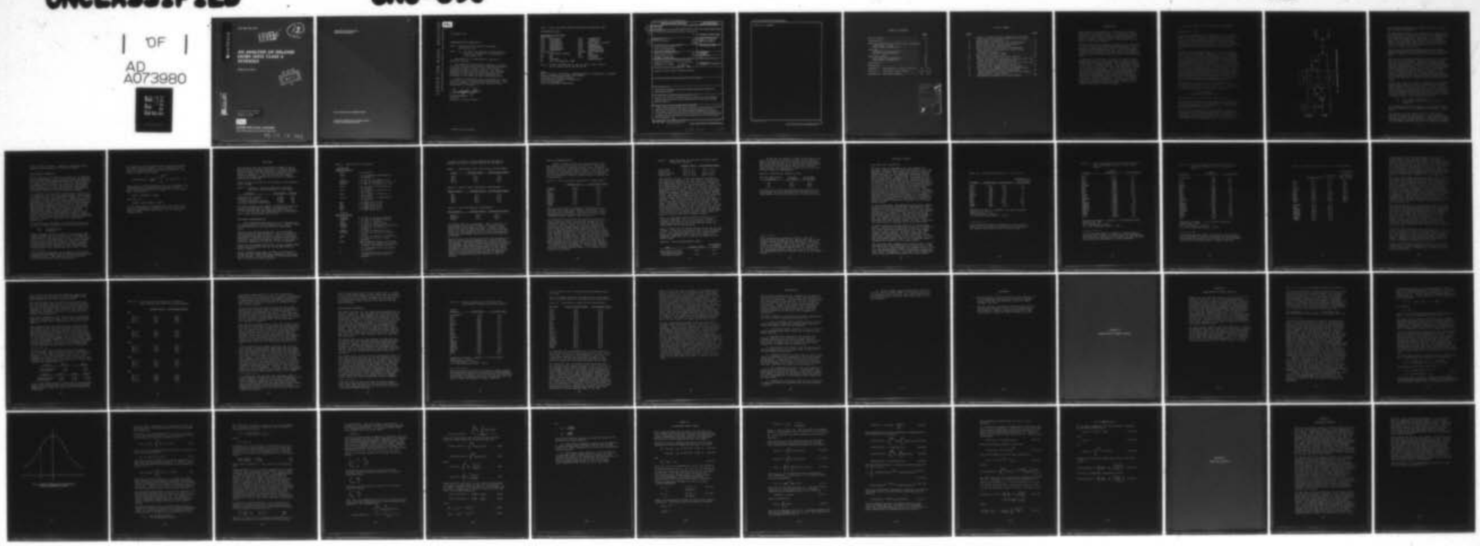
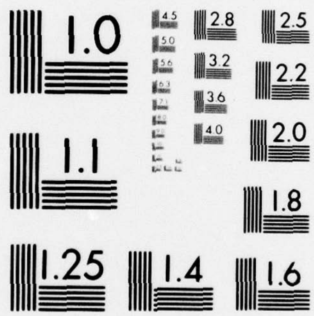
AD-A073 980 CENTER FOR NAVAL ANALYSES ALEXANDRIA VA F/G 5/9
AN ANALYSIS OF DELAYED ENTRY INTO CLASS A SCHOOLS.(U)
MAY 79 W M EVANCO N00014-76-C-0001
UNCLASSIFIED CRC-390 NL

1 OF 1
AD
A073980



END
DATE
FILMED
10-79
DDC



MICROCOPY RESOLUTION TEST CHART

NATIONAL BUREAU OF STANDARDS-1963-A

AD A073980

CRC 390 / May 1979

LEVEL II

12

B.S.

AN ANALYSIS OF DELAYED ENTRY INTO CLASS A SCHOOLS

William M. Evanco

CNA

CENTER FOR NAVAL ANALYSES

2000 North Beauregard Street, Alexandria, Virginia 22311

79 09 18 005

Work conducted under contract N00014-76-C-0001

CENTER FOR NAVAL ANALYSES
2000 North Beauregard Street, Alexandria, Virginia 22311 Telephone: (703) 998-3500

CNA

16 August 1979

MEMORANDUM FOR DISTRIBUTION

Subj: Center for Naval Analyses Research Contribution 390

Encl: (1) CRC 390, "An Analysis of Delayed Entry into Class A Schools," by William M. Evanco, May 1979

1. Enclosure (1) is forwarded as a matter of possible interest.

2. This Research Contribution, prepared in connection with the Enlisted Accession and Training Study, evaluates delayed A School entry versus entry immediately after recruit training. The results should be useful for both recruiters and A School planners in selecting training strategies for enlistees on the basis of their characteristics.

3. Research Contributions are distributed for their potential value in other studies and analyses. They do not necessarily represent the opinion of the Department of the Navy.

Christopher Jehn

CHRISTOPHER JEHN
Director
Institute of Naval Studies

Subj: Center for Naval Analyses Research Contribution 390

DISTRIBUTION LIST

Department of the Navy

SNDL Part I:

21A1	CINCLANTFLT	FF30	NAVMMACLANT
21A2	CINCPACFLT	FF30	NAVMMACPAC
21A3	CINCUSNAVEUR	FF38	USNA, Nimitz Library
22A1	COMSECONDFLT	FF44	NAVWARCOL
22A2	COMTHIRDFLT	FF48	HUMRESMANCEN
22A2	COMSEVENTHFLT	FJ18	NAVMILPERSCOM
22A3	COMSIXTHFLT	FJ76	COMNAVCRUITCOM
SNDL Part II:		FKA6A16	NAVPERSRANDCEN
A1	ASST SECNAV (MRA&L)	FT1	CNET
A2A	OPA	FT5	CNTECHTRA
A2A	CNR	FT73	NAVPGSCOL
A5	CHNAVPERS		NAVTACSUPPACT
A6	DC/S for Manpower, HqMC		

OpNav: Op-00K, Op-09BH, Op-01, Op-11, Op-12, Op-13, Op-14, Op-15, Op-29, Op-39, Op-49, Op-59

Other

Ass't Secretary of Defense, Program Analysis & Evaluation (2 copies)
Defense Documentation Center (12 copies)
Institute for Defense Analyses
Human Resource Research Organization
The Rand Corporation
System Development Corporation

SECURITY CLASSIFICATION OF THIS PAGE (When Data Entered)

REPORT DOCUMENTATION PAGE

READ INSTRUCTIONS BEFORE COMPLETING FORM

14

1. REPORT NUMBER: CRC-390

2. GOVT ACCESSION NO.

3. RECIPIENT'S CATALOG NUMBER

6

4. TITLE (and Subtitle): An Analysis of Delayed Entry into Class A Schools

5. TYPE OF REPORT & PERIOD COVERED

6. PERFORMING ORG. REPORT NUMBER

10

7. AUTHOR(s): William M. Evanco

15

8. CONTRACT OR GRANT NUMBER(s): N00014-76-C-0001

9. PERFORMING ORGANIZATION NAME AND ADDRESS
Center for Naval Analyses
2000 N. Beauregard Street
Alexandria, Virginia 22311

10. PROGRAM ELEMENT, PROJECT, TASK AREA & WORK UNIT NUMBERS

11. CONTROLLING OFFICE NAME AND ADDRESS
Office of Naval Research
Department of the Navy
Arlington, Virginia 22217

11

12. REPORT DATE: May 1979

13. NUMBER OF PAGES: 59

14. MONITORING AGENCY NAME & ADDRESS (if different from Controlling Office)
Office of the Chief of Naval Operations (Op 96)
Department of the Navy
Washington, D.C. 20350

12 52p.

15. SECURITY CLASS. (of this report): Unclassified

15a. DECLASSIFICATION/DOWNGRADING SCHEDULE

16. DISTRIBUTION STATEMENT (of this Report)

Approved for public release; distribution unlimited.

17. DISTRIBUTION STATEMENT (of the abstract entered in Block 20, if different from Report)

18. SUPPLEMENTARY NOTES

This Research Contribution does not necessarily represent the opinion of the Department of the Navy.

19. KEY WORDS (Continue on reverse side if necessary and identify by block number)

attrition, class A schools, enlisted personnel, measurement, Naval personnel, Naval training, performance (human), probit analysis, reenlistment, retention (general), schools

20. ABSTRACT (Continue on reverse side if necessary and identify by block number)

This report evaluates delayed Class A technical school entry versus immediate entry after recruit training. The effects of recruit background and service characteristics on first-term loss and reenlistment probabilities are measured. Conclusions are drawn about the selection of enlistees for A school attendance either immediately after recruit training or after general duty in the fleet on the basis of their characteristics.

DD FORM 1 JAN 73 1473 EDITION OF 1 NOV 65 IS OBSOLETE
S/N 0102-LF-014-6601

SECURITY CLASSIFICATION OF THIS PAGE (When Data Entered)

077 270

gw

[7] Bruce Simon, Contributor

TABLE OF CONTENTS

LIST OF TABLES

INTRODUCTION

Most enlistees attending Class A schools do so with guarantees of Navy schooling. They qualify as school eligible on the basis of various combinations of test scores from the Armed Services Vocational Aptitude Battery (ASVAB). Some enlistees later qualify on the basis of test score waivers.

The difficulty with guaranteed schooling upon enlistment is that the selection process depends heavily on the results of paper-and-pencil tests. While these tests measure skills and aptitudes that affect the chances of successfully completing training, they do not measure personal characteristics that may affect success in Navy jobs.

In addition to selecting enlistees for Class A schools immediately after recruit training, the Navy selects individuals for Class A schools after they have served in the fleet on general detail. These delayed A school entries may or may not have been qualified for training at the time of enlistment. With such enlistees, the Navy has had the opportunity for additional screening and evaluation before making a training commitment. On-the-job performance and personal characteristics, in addition to background data collected at the time of enlistment, can be considered. Further, an enlistee with fleet experience may have more realistic career expectations when he enters A school, resulting in a lower probability of early discharge.

MODEL FOR ANALYSIS OF A SCHOOL DELAY STRATEGIES

PROBABILITY OF LOSS

To study the advantages and disadvantages of delayed A school training, we ought to compare measures of performance for enlistees who had fleet experience before starting A school with those who had attended school immediately after recruit training. Moreover to make a fair comparison other factors affecting performance must be held constant by either empirical or statistical means.

Ideally, the performance measures should be directly related to an individual's productivity or performance on the job; such measures unfortunately are not available. However, as discussed in appendix A, the probability that an enlistee does not complete his first term (loss probability) can be viewed as a crude measure of job performance in the following sense: If the benefits received by the Navy are less than the costs for a given enlistee, then he ought to be discharged before the end of his active obligated service. Thus, whether or not an individual is prematurely discharged should be determined by his location on a continuous scale of performance relative to cost; below a certain threshold value he is terminated, above the threshold he is not. We will use the four-year loss probability as a proxy for a more exact, but unavailable, performance measure.

The model for the analysis of loss probabilities is illustrated in figure 1. Each individual in the cohort is characterized with respect to his loss or survival at the end of four years by a variable L defined as

$$L = \begin{cases} 1 & \text{if early discharge} \\ 0 & \text{if survive for four years} \end{cases}$$

depicted on the far right hand side of figure 1.

Benefits derived from a recruit depend upon productivity, which is a function of characteristics such as motivation, adaptability, and perseverance, in addition to aptitudes and relevant skills.

Costs involve military pay and allowances, training, transportation, subsidies, health care, counseling and legal aid, and various indirect costs of personnel management. Some of these costs, such as military pay, are relatively constant among first term enlistees,

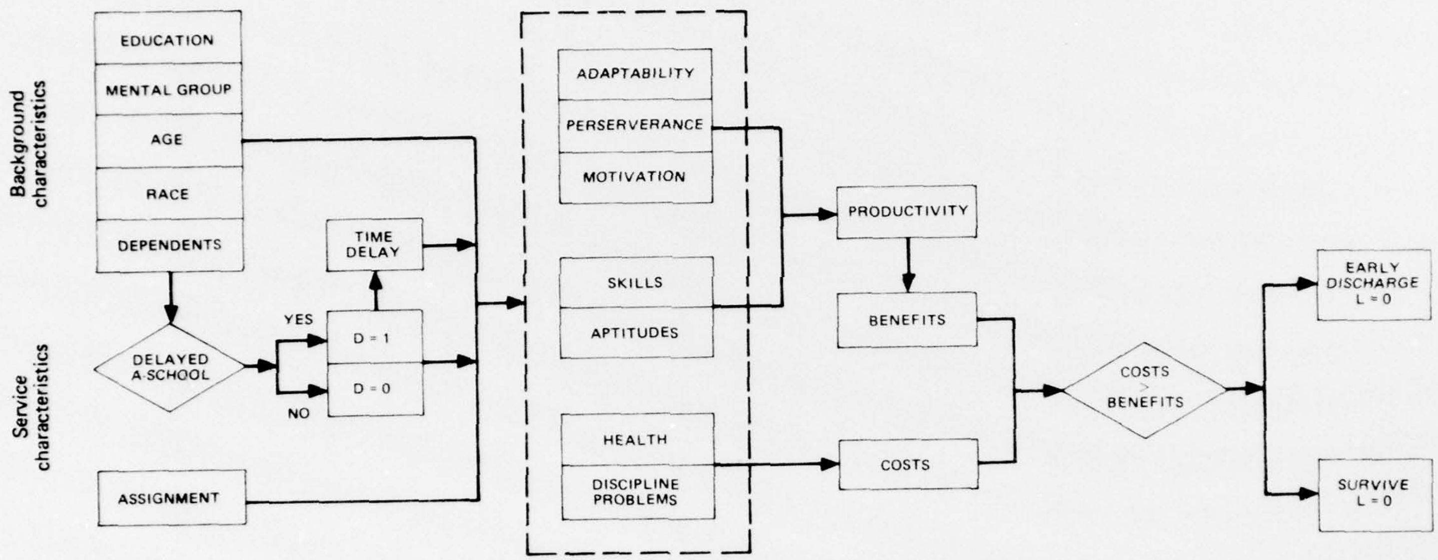


FIG. 1: SCHEMATIC MODEL FOR LOSS PROBABILITIES

while health and disciplinary costs will vary greatly for different individuals.

The attributes listed in the dashed box of figure 1 have not been directly measured, but they can be proxied by the background and service characteristics (on the far left) that are empirically available. For example, educational attainment can be taken to be a rough indicator of perseverance, motivation, and skills that the recruit brings to the Navy, and mental group can be taken as a proxy for aptitude.

Disciplinary and health problems may depend on background characteristics and assignment, since living and working conditions are expected to differ by assignment.

The timing of A school training is crucial to this analysis for several reasons. For those who were school-delayed, the Navy had the opportunity to make evaluations and selections on the basis of job performance and personal characteristics. All else being equal, school-delayed individuals ought to have lower loss probabilities. Moreover, school-delays have a shorter time period over which they can be lost compared to non-delayed A school graduates. To make fair comparisons of loss probabilities for delayed and non-delayed trainees, it will be necessary to control statistically for time delay.

Lastly, note that in figure 1 a dependence is indicated between the timing of A school training and the background characteristics of the recruit. Some persons of lower quality (as measured by educational attainment or mental group) but having desirable personal characteristics (as deduced from job performance) will be selected to enter A school after having served on general detail. A dummy variable D can be defined such that

$$D = \begin{cases} 1 & \text{if A school is delayed} \\ 0 & \text{if A school immediately after recruit training} \end{cases}$$

The probability of school delay as a function of background variables can be analyzed by using probit analysis.

The analysis of loss probabilities is complicated because the data was generated by a selection process based on personal characteristics. The problem is that the process of selection produces an empirical distribution of losses that does not meet the normality require-

ment of probit analysis. Appendix A describes probit analysis and the problem of selectivity bias.

REENLISTMENT PROBABILITY

Another measure of interest to the Navy is the probability of reenlistment. All else being equal, an enlistee with a higher reenlistment probability is preferred over one with a lower probability. Since the time between the completion of A school and the end of active obligated service is less for school-delayed individuals, higher reenlistment rates are needed to compensate for the reduced amount of trained services received.

The formalism for estimating reenlistment probabilities is similar to that for loss probabilities. Reenlistment is conditional upon the successful completion of a four-year tour of duty. Whether or not a person reenlists is presumably a function of his background characteristics and the type of Navy activity to which he was assigned. Employment and schooling opportunities in the civilian sector differ depending on factors such as education, mental group, race, age, etc.; greater civilian opportunities imply a lower tendency to reenlist. On the other hand, those with dependents might be risk averse, and hence, more willing than others to reenlist. Finally, the activity type determines working conditions which are expected to influence the reenlistment probability.

For the cohort of enlistees who survived four years in the Navy, a dummy variable R can be defined such that

$$R = \begin{cases} 1 & \text{if reenlistment} \\ 0 & \text{otherwise} \end{cases}$$

Probit analysis can then be used to estimate the reenlistment probability as a function of the background characteristics and the activity type. As in the calculation of the loss probability, we might expect selectivity bias to enter the analysis. However, initial analyses showed that the coefficients associated with the selectivity bias corrections were statistically insignificant and the corrections were small.[1] Thus,

---

[1] An alternative approach was attempted by pooling the delayed and non-delayed cohorts, and then accounting for selectivity bias. Once again the selectivity bias corrections were small and insignificant.

we decided to use the pooled cohort approach including the school delay variable, D, as an intercept shift term, but ignoring selectivity bias corrections. The probit integral then becomes:

$$\text{Prob}(R=1|\,X,D) = \left(\frac{1}{2\pi}\right)^{1/2} \int_{-\infty}^{X\beta+\omega D} du \; \exp(-u^2/2) \qquad (4)$$

where β and ω are the coefficients to be estimated. The introduction of the explanatory variable D serves to shift the constant term in the upper limit of the integral. Thus for D=0

$$X\beta+\omega D = \beta_0+\beta_1 X_1+\ldots+\beta_n X_n$$

while for D=1

$$X\beta+\omega D = (\beta_0+\omega)+\beta_1 X_1+\ldots+\beta_n X_n$$

If the coefficient ω is greater than (less than) zero, then the reenlistment probability for the school delay group is greater than (less than) that for the non-delayed group.

DATA BASE

The four-year loss and reenlistment probabilities of delayed and non-delayed completors of Class A schools were estimated from service data for calendar year 1973 entrants; details of the data base creation are presented in appendix C. The 1973 cohort was chosen because the cohort members have had time to complete four years of service.

Table 1 gives the sizes of the delayed and non-delayed A school groups.

Table 1: Numbers of Class A Completers, Four-Year Survivors, and Reenlistees in 1973 Cohort

Category	Non-delayed	Delayed
Completers of A school	33,850	876
Four-year survivors	25,470	613
of which four-year obligors	18,061	611
Four-year obligors reenlisting	3,849	252

For each enlistee in the sample, background and service history data defined in table 2 was extracted; all variables except TDEL are dummy variables (values of 0 or 1). The activity type variables refer to an individual's first non-training tour of duty.

BACKGROUND CHARACTERISTICS

The background characteristics of the delayed and non-delayed cohorts differ substantially. Table 3 gives the distribution of the two groups by educational level.

Persons in the delayed cohort have, on the average, lower educational qualifications. About 40 percent of the delayed group had less than 12 years of education compared to 15 percent of the non-delayed group, while 60 percent among the delayed had twelve or more years of education compared to 85 percent of the non-delayed.

Similarly, the delayed cohort had, on the average, lower mental qualifications than the non-delayed group as shown in table 4.

School eligibles (MG1, MG2, and MG3U) constituted 85 percent of the non-delayed group compared to 40 percent of the delayed. Thus, 60 percent of the delayed group were not initially school eligible.

Table 2: Definitions of Variables

Background characteristics	
RACE	1 if nonwhite
PDEPS	1 if any primary dependents at enlistment
AGE17	1 if age at enlistment is 17
AGE18/19	1 if age at enlistment is 18 or 19
AGE20P	1 if age at enlistment is $\geq$ 20
ED<11	1 if <11 years of education at enlistment
ED11	1 if 11 years of education at enlistment
ED12	1 if 12 years of education at enlistment
ED>12	1 if >12 years of education at enlistment
MG1	1 if AFQT score 95-100
MG2	1 if AFQT score 67-94
MG3U	1 if AFQT score 50-66
MG3L	1 if AFQT score 34-49
MG4	1 if AFQT score 21-33
Service characteristics	
SURFACE	1 if duty on surface combatant
CARRIER	1 if duty on aircraft carrier
SUB	1 if duty on submarine
REPAIR	1 if duty on repair vessel
SBSAIR	1 if duty in sea-based air squadron
LBSAIR	1 if duty in land-based air squadron
AMPHIB	1 if duty on amphibious ship
AUX/PTL	1 if duty on auxiliary or patrol vessel
SBEE	1 if duty in construction battalion
D	1 if entry to Class A school delayed
TDEL	Time in months between active duty base date and completion of first A school (continuous variable)
L	1 if discharged before four years of service
R	1 if reenlisted after successful completion of four years of service

As shown in table 5, the delayed group contained a greater fraction of younger persons at the time of

Table 3: Educational Level Distribution (Percentages)

Level	Delayed cohort	Non-delayed cohort
ED<11	23.1	6.8
ED11	17.2	7.9
ED12	55.9	74.2
ED>12	3.8	11.1

Table 4: Mental Group Distribution (Percentages)

Mental group	Delayed cohort	Non-delayed cohort
MG1	0.5	7.5
MG2	11.1	53.1
MG3U	28.1	24.4
MG3L	36.4	11.0
MG4	23.9	4.0

Table 5: Age Distributions (Percentages)

Age category	Delayed cohort	Non-delayed cohort
AGE17	35.8	21.0
AGE18/19	50.3	60.4
AGE20P	13.9	18.6

enlistment than the non-delayed. About 36 percent of the delayed cohort were 17 years old compared to 21 percent of the non-delayed cohort. Previous studies have shown that 17 year olds tend to exhibit high loss rates, so that postponing their further training until on-the-job performance data can be evaluated might be warranted.

About 6 percent of the delayed cohort had dependents at the time of enlistment compared to 7 percent for the non-delayed. Minorities constituted about 19 percent of the delayed group compared with only 6 percent for the non-delayed. Relatively fewer minorities are school eligible by mental group qualification at the time of enlistment; thus the screening of such persons on the basis of job performance could provide an important source of successful minority entrants to Class A schools.

SERVICE CHARACTERISTICS

Table 6 shows the activity distributions of the two cohorts for their first tour of duty. Note that about 24 percent of delayed entrants first served on aircraft carriers, while only 16 percent of recruits are assigned to carriers on general detail. This may be due to the selection process for carrier duty, or because such duty provides superior opportunities to qualify for delayed entry.

Table 6: First Duty Tour Distribution (Percentages)

	Delayed cohort	Non-delayed cohort
SURFACE	21.0%	19.0%
CARRIER	24.3	9.3
SUB	1.4	5.0
REPAIR	8.0	6.9
SBSAIR	11.6	9.5
LBSAIR	2.1	5.7
AMPHIB	10.8	5.7
AUX/PTL	9.1	5.6
SBEES	1.3	2.7
OTHER	10.3	30.6

Note also that about 10 percent of delayed A school entrants are listed under "OTHER" versus about 31 percent for the non-delayed. This group contains those individuals who do not fit into any of the duty types "SURFACE" through "SBEES" shown in table 2.

The analysis of alternative strategies for delayed A school training cannot be performed adequately without information about the timing of the delays and their effects on service times. In table 7, "Time Delay" represents the average number of months between the active duty base date and the completion date of the first A school attended (the plus or minus terms represent standard deviations). For the non-delayed cohort, "Time Delay" includes recruit training time and A school time; for the delayed cohort, "Time Delay" also includes "Fleet Time", defined as the difference between the starting date of the first A school attended and the date received into the first duty tour. The average "Time Delay" was 18.8 months for the delayed cohort and 5.3 months for the non-delayed. "Fleet Time", during which the delayed training cohort is on general detail, averaged 12.3 months.

Table 7: Times Relevant for Analysis of School Delay Strategies (Months)

	Delayed cohort	Non-delayed cohort
Time delay	18.8 ± 8.3	5.3 ± 2.2
Fleet time	12.3 ± 8.1	---- ----
Payback period	25.3 ± 11.2	36.9 ± 12.2
Total time	37.6 ± 11.41	36.9 ± 12.2

The "Payback Period" -- the time between the date of attrition (or completion of four years of service) and the completion date for the first A school attended -- was appreciably lower for the delayed group. This time is a rough measure of the amount of skilled services the Navy receives from a four-year first termer after his graduation from A school. Since an enlistee may attend additional schools throughout his Navy career, "Payback Period" represents an upper bound on the amount of services. The average "Payback Period" was 25.3 months for school-delayed persons and 36.9 months for the non-delayed. This 12 month difference is due principally to the "Fleet Time" served by the school-delayed. "Total Time", reflecting the total amount of services to the Navy, is slightly higher for the delayed group. For the non-delayed cohort, "Total Time" is defined as the time from the date received into the first fleet assignment until the date of attrition (or completion of four years of service); for the delayed cohort, "Total Time" is equal to "Fleet Time" plus "Payback Period".

Thus, the Navy gets lower first-term returns on its training investment for enlistees who are school-delayed. This effect will be mitigated, however, if reenlistment rates for the school-delayed group are higher than for the non-delayed.

Table 8 shows that the school-delayed group had, on the average, twice the reenlistment rate of the non-delayed group. The table also shows that the school-delayed group had a somewhat lower loss rate, despite its generally lower quality as measured by education and mental group.

Table 8: Loss and Reenlistment Rates

Rate	Delayed cohort	Non-delayed cohort
Four-year loss rate	30.0	37.1
Reenlistment rate for four-year obligors	41.2	21.3

To see the net effect of these two factors on the flow of services to the Navy, we add to the "Service Time" the expected months of service from reenlistment, defined as the product of the reenlistment probability and the term of reenlistment assuming no attrition.[1] These results are tabulated in table 9 for 24, 36, and 48 month terms of reenlistment.

Table 9: Net Flow of Service to Navy

Term of reenlistment (in months)	Delayed (in months)	Non-delayed (in months)
24	35.1	41.9
36	40.1	44.5
48	45.0	47.0

In every case, the school delayed group provides less post-school services, although for a 48-month term of reenlistment the difference reduces to only two months.

[1]This calculation is biased in favor of the non-delayed cohort because we have ignored the flow of services generated by delayed recruits before entering A school. In addition, the total time spent in training (including multiple A school attendance, C schools, etc.) by the non-delayed cohort might be greater than for the delayed cohort; thus, the flow of services of the non-delayed would be reduced relative to the delayed cohort.

EMPIRICAL RESULTS

FOUR YEAR LOSS PROBABILITY

The first step in estimating the loss probability involves a probit analysis with school delay, D, as the dependent variable. The results are shown in table 10. Coefficients of all the independent variables except AGE20P and PDEPS are statistically significant at the five percent level. Equation (A6) of appendix A was used to compute the probabilities. The characteristics of persons in the base case are specified in footnote (a) of table 10 and are called the base characteristics; for each of the other variables, the probability shown is for the characteristic defined by that variable substituted for its counterpart in the base. Thus, a recruit with all the base characteristics has a school delay probability of 0.4 percent; if he is in MG4 rather than MG2, this probability increases to 9.4 percent. Consistent with our aggregate analysis of the previous section, the probability of school delay is higher for those with lower educational attainment and for lower mental groups.

In the second stage, the background and service characteristics along with variable W_1, representing a correction for the selection process, are used in a probit analysis to find the probability of loss. These results are presented in table 11. The coefficients of the background characteristic variables are significant at the five percent level with the exception of AGE20P, MG1, and PDEPS, while the only significant service variables are SUB, AUX/PTL, OTHER, and TDEL. Note that the coefficient of variable W_0 is also significant.

A similar analysis has been made for the non-delayed cohort. The results of which are presented in table 12. The coefficients are significant at the five percent level with the exception of AGE2OP, MG3L, MG4, RACE, CARRIER, REPAIR, and W_0. The small magnitude and statistical insignificance of the coefficient for W_0 imply that selectivity bias plays a very small role for the non-delayed cohort; this is to be expected since it comprises 97.5 percent of the total cohort.

The four-year loss probabilities shown in table 13 were calculated from the coefficients of tables 10, 11, and 12. The first column presents the probabilities of loss from the delayed school cohort estimated by using equation (A10a) in Appendix A. The time, TDEL, between the active duty base date and completion of the first A

Table 10: School Delay Probability (Total cohort)

Variable	Coefficient	t-statistic	Probability of school delay (in percent)
BASE[a]	-2.63	66.9	.4
ED<11	.429	8.9	1.4
ED11	.253	5.3	.9
ED>12	-.164	2.0	.3
AGE17	.091	2.3	.6
AGE20P	.079	1.5	.5
MG1	-.329	2.1	.2
MG3U	.580	12.6	2.0
MG3L	1.04	22.1	5.7
MG4	1.31	23.0	9.4
PDEPS	.009	.13	.4
RACE	.157	3.2	.7

Dependent Variable: D (equals 1 if school delayed)
Sample Size: 34,726
Log(Likelihood) = -3413.1
-2 x Log(Likelihood Ratio) = 1365.4

---

[a] The base group refers to Caucasian, mental group 2 recruits with 12 years of education and no dependents, who entered the Navy at 18 or 19 years of age.

Table 11: Probit Estimates of Coefficients for Four-Year Loss Probabilities of School-Delay Cohort

Variable	Probit coefficient	t-statistic
Base[a]	-2.905	2.1
ED<11	.3715	2.2
ED11	.2268	2.3
ED>12	-.1612	2.4
AGE17	.0899	2.5
AGE20P	.0597	1.8
MG1	-.2593	1.7
MG3U	.5129	2.1
MG3L	.9192	2.2
MG4	1.1273	2.1
PDEPS	.0148	1.1
RACE	.1349	2.2
CARRIER	-.0048	.52
SUB	-.0062	1.9
REPAIR	.0006	.05
SBSAIR	-.0172	1.5
LBSAIR	-.0395	1.6
AMPHIB	.0122	1.1
AUX/PTL	-.0253	2.0
SBEE	-.0417	1.2
OTHER	.0242	2.1
TDEL	-.0032	7.3
W	.9978	2.2

Dependent Variable: L (equals 1 if discharged early)
Sample size: 876
Log (Likelihood) = -480.6
-2 x Log (Likelihood ratio) = 109.3

---

[a] The base term refers to Caucasian, mental group 2 recruits with 12 years of education and no dependents, who entered the Navy at 18 or 19 years of age, and whose first duty tour was aboard a surface combatant.

Table 12: Probit Estimates of Coefficients for Four-Year Loss Probabilities of Non-Delayed Cohort

Variable	Probit coefficient	|t-statistic|
Intercept[a]	-.344	15.3
ED<11	.440	9.3
ED11	.345	10.9
ED>12	-.160	5.9
AGE17	.159	8.1
AGE20P	-.014	.6
MG1	-.125	4.3
MG3U	.068	2.5
MG3L	.109	1.6
MG4	.088	.8
PDEPS	.065	2.2
RACE	-.017	.5
CARRIER	.018	.7
SUB	-.262	7.1
REPAIR	-.053	1.7
SBSAIR	-.157	5.5
LBSAIR	-.215	6.2
AMPHIB	.138	4.2
AUX/PTL	.082	2.4
SBEE	.169	3.8
OTHER	.132	6.4
TDEL	-.020	7.8
W_0	.083	.2

Dependent Variable: L (=1 if early discharge)
Sample Size: 33850
Log (Likelihood) = -21650.0
-2 x Log (Likelihood ratio) = 1347.9

[a]The intercept term refers to Caucasian, mental group 2 recruits with 12 years of education and no dependents, who entered the Navy at 18 or 19 years of age, and whose first tour of duty was aboard a surface combatant.

Table 13: Four-Year Loss Probabilities (Percentages)

	Delayed cohort	Non-delayed cohort	School delayed cohort for activity type "Other"
BASE	30.9	32.5	44.7
ED<11	22.3	49.5	34.6
ED11	29.3	45.7	42.9
ED>12	25.3	26.9	38.2
AGE17	34.5	38.3	48.7
AGE20P	24.6	32.0	37.4
MG1	56.1	28.1	69.8
MG3U	25.8	35.0	38.9
MG3L	28.2	36.5	41.6
MG4	18.8	35.7	30.1
PDEPS	34.6	34.8	48.8
RACE	26.8	31.9	40.0
CARRIER	28.4	33.1	
SUB	27.7	23.7	
REPAIR	31.2	30.6	
SBSAIR	22.4	27.1	
LBSAIR	13.7	25.1	
AMPHIB	37.6	37.6	
AUX/PTL	18.9	35.5	
SBEES	13.0	38.8	
OTHER	44.7	37.4	

school was set equal to 18.84 months, which was the average TDEL for the delayed cohort.[1] The base characteristics for the first column are those given in footnote (a) of table 11; for each of the other variables, the probability indicated is for that characteristic replacing its counterpart in the base. Thus, a school-delayed enlistee with the base characteristics has an estimated loss probability of 30.9 percent, while one with the same base characteristics, except for membership in MG3U, has a probability of 25.8 percent.

Note that the delayed enlistees of lower quality tend to have lower loss probabilities than those with better qualifications who are not delayed. The probability of loss for a school-delayed enlistee in MG4 is about 19 percent compared to about 31 percent for MG2 recruits and 56 percent for MG1 recruits.[2] An enlistee with less than 11 years of education has a loss probability of about 22 percent, compared to 31 percent for one with 12 years and 25 percent for one with more than 12 years of education. These results are consistent with the interpretation that persons of lower quality are screened more carefully through job performance than are those who meet the Navy's primary requirements for school eligibility.

The loss probabilities in the second column are for the non-delayed cohort, the base characteristics being the same as for the school-delayed group. These probabilities were estimated from equation (A11a) in appendix A; the delay time, TDEL, was set equal to 5.32 months, which is the average time elapsed between the active duty base date and A school completion for the non-delayed cohort. In contrast to the school-delayed cohort, loss probabilities increase as educational attainment and mental group standing decrease. Moreover, with the exception of MG1 recruits, the probabili-

---

[1]A similar calculation was done with TDEL set to 12 months resulting in about a 13 percentage point increase in each probability in the first two columns of table 13. The value of 12 months for TDEL implies approximately 6 months of fleet experience before entering A school. Apparently, such a short fleet time does not allow for adequate screening of the recruit.

[2] The result for MG1 recruits is based on a very small sample and on a probit coefficient estimate significant to only 10 percent. Consequently, it should be interpreted with caution.

ties of loss are consistently higher for each of the background characteristics of the non-delayed cohort when compared to the school-delayed cohort.

The loss probabilities for various mental group and educational level combinations are presented in table 14. The probabilities are for those mental group and educational characteristics substituted for their counterparts in the base. Thus, an MG3L, ED11 enlistee whose other characteristics are in the base has a 47.3 percent loss probability.

Note that, compared to their non-delayed counterparts, delayed enlistees with (MG2, ED<11), (MG2, ED11), (MG4, ED12), and (MG4, ED>12) characteristics perform very well.[1]

When comparing the loss probabilities associated with the activity types for the two groups in table 13, we get mixed results. However, note that the activity types enter the probit analyses and, hence, the probability estimates for the two cohorts in different ways. For the school-delayed cohort, the activity type refers to the first duty tour before completing A school, while for the non-delayed group the activity type occurs after A school completion. Thus, for a non-delayed recruit, the activity type is an indicator of the working and living conditions he must face, which are expected to

---

[1] Although lower quality school delayed individuals tend to exhibit lower loss probabilities, the possibility exists that they have higher dropout rates from A schools. If so, then recommendations based on loss probabilities alone will not be valid. To examine this possibility, dropout rates are tabulated by mental group standing and educational level:

Cohort	MG1-3U	MG3L-4
School-delayed	11%	3.7%
Non-delayed	9.4%	24.0%

Cohort	ED<11	ED12	ED>12
School-delayed	4.9%	8.5%	14.3%
Non-delayed	21.0%	18.0%	6.3%

We see that lower quality enlistees who are delayed have lower dropout rates, while the non-delayed have higher dropout rates on both the mental group and educational dimensions.

Table 14: Four-Year Loss Probabilities for Mental Group, Educational Combinations (Percentages)

	Delayed cohort	Non-delayed cohort
MG1		
ED < 11	57.1	44.8
ED 11	67.1	40.9
ED 12	56.1	28.1
ED > 12	66.4	22.9
MG2		
ED < 11	22.3	49.5
ED 11	29.3	45.7
ED 12	30.9	32.5
ED > 12	25.3	26.9
MG3U		
ED < 11	36.1	52.6
ED 11	40.9	48.7
ED 12	25.8	35.0
ED > 12	30.3	29.4
MG3L		
ED < 11	46.5	54.5
ED 11	47.3	50.5
ED 12	28.2	36.5
ED > 12	28.8	31.0
MG4		
ED < 11	42.2	53.9
ED 11	39.9	50.0
ED 12	18.8	35.9
ED > 12	18.4	30.5

contribute differentially to the loss probability. Additionally, the activity type may reflect differential characteristics of enlistees; for example, higher quality individuals and those with desirable personality characteristics might be assigned to submarines rather than surface vessels.

On the other hand, the activity type for the school-delayed cohort results from a screening process providing differential opportunities for an individual to demonstrate his aptitudes and skills. For those individuals who return to the same activity type after completing A school, this variable would also be an indicator of differential working and living conditions contributing to losses.[1]

With the above distinctions in mind, and further noting from table 11 that most of the coefficients associated with the activity type variables are not significant at the five percent level, we see that the school-delayed cohort performs about as well as or better than the non-delayed for all activity types except "SUB" and "OTHER". The school-delayed enlistees in land-based air and construction battalion activity types have exceptionally low loss rate estimates, but the statistical significance levels of the associated coefficients is low.

A loss probability of almost 45 percent for activity type "OTHER" is appreciably higher than the loss probabilities associated with the specified activity types. The third column of table 13 shows some loss probabilities associated with each of the background characteristics for a recruit whose first duty tour was in activity type "OTHER". A possible explanation for these high loss probabilities is that individuals in the "OTHER" category have not been adequately screened. This activity type usually involves shore duty and might provide inadequate information about a recruit's abilities to adapt to a fleet assignment. Moreover, such individuals might have been assigned to an A school while in recruit training and then spent several months in the "OTHER"

---

[1] In principle, the activity type variables after A school completion could have also been introduced in the probit analysis. However, the introduction of additional variables would tax our relatively small data base and multicollinearity problems would further reduce significance levels. About 65 percent of school-delayed recruits return to the same activity type after completing A school.

activity type while waiting for a school seat. In support of this proposition, we note that about 45 percent of recruits in the "OTHER" category complete their first A school within nine months of their active duty base date compared to about 9 percent for the rest of the activity types.

REENLISTMENT PROBABILITY

The probit estimate for the reenlistment probability is shown in table 14. The variables are identical to those used to explain the loss probability, but with several exceptions. Since reenlistments are conditional on the successful completion of four years of service, the time delay factor, TDEL, is no longer relevant to the analysis. The school delay variable, D, is introduced to differentiate the delayed from the non-delayed group, because the probit analysis is applied to the total cohort. Finally, the activity type variables, CARRIER through OTHER, refer to an individual's second tour (or the first if he had only one tour), since it is during this tour that he is most likely to make the reenlistment decision.[1]

The cohort used for the reenlistment analysis consists of 18,672 four year obligors who have successfuly completed four years of service; any individual who had a nuclear or advanced electronics field indicator was assumed to be a six-year obligor and excluded from the cohort. The base characteristics, specified in footnote (a) of table 14, are the same as those in the loss analysis.

The coefficients in table 15 have significance levels of less than five percent for all of the background variables except ED>12 and MG1, and all of the service variable coefficients except REPAIR and AUX/PTL. Note that the positive coefficient of the school delay variable, D, implies that delayed enlistees have a higher reenlistment probability than the non-delayed. Also, the magnitudes of the coefficients associated with the service characteristics, except for PDEPS and RACE, tend to be larger than those of the coefficients of the background characteristics; thus educational attainment, mental group, and age are expected to play a compara-

---

[1]Very few recruits have more than two tours during their four- year service, so the second tour reflects with reasonable accuracy the activity type at the time of the reenlistment decision.

Table 15: Probit Estimates of Coefficients for Reenlistment Probabilities (Total Cohort)

Probit variable	Coefficient	t-statistic
Intercept[a]	-1.07	3.6
ED<11	.110	2.4
ED11	.076	1.9
ED>12	-.049	1.2
AGE17	.083	2.8
AGE20P	.070	2.5
MG1	-.086	1.7
MG3U	.069	2.7
MG3L	.090	2.7
MG4	.135	2.7
PDEPS	.326	7.8
RACE	.294	7.2
CARRIER	-.199	4.3
SUB	.523	9.1
REPAIR	.024	.5
SBSAIR	.103	2.5
LBSAIR	.355	7.5
AMPHIB	-.122	2.4
AUX/PTL	-.075	1.4
SBEE	-.350	4.8
OTHER	.431	13.4
D	.512	9.3

Dependent Variable: R(equals 1 if reenlistment)
Sample Size: 18,672
Log(Likelihood) = -9371.6
-2 x LOG(Likelihood ratio) = 916.4

[a]The intercept term refers to Caucasian, mental group 2 enlistees with 12 years of education and no dependents, who entered the Navy at 18 or 19 years of age, and whose second tour of duty(or first tour if there was only one tour of duty) was aboard a surface combatant.

tively smaller role in determining reenlistment probabilities.

Table 16 shows reenlistment probabilities for delayed and non-delayed cohorts. The base case probabilities of

Table 16: Reenlistment Probabilities (Percentages)

Variable	School delayed cohort	Non-delayed cohort
BASE	28.7	14.2
ED<11	32.6	16.8
ED11	31.4	15.9
ED>12	27.1	13.1
AGE17	31.6	16.1
AGE20P	31.5	16.0
MG1	25.9	12.3
MG3U	31.1	15.8
MG3L	31.9	16.3
MG4	33.5	17.4
PDEPS	40.7	22.8
RACE	39.5	21.8
CARRIER	22.4	10.2
SUB	48.5	29.1
REPAIR	28.8	14.7
SBSAIR	32.3	16.6
LBSAIR	41.8	23.6
AMPHIB	24.7	11.6
AUX/PTL	26.2	12.5
SBEES	18.1	7.7
OTHER	44.8	26.0

28.7 percent for the delayed and 14.2 percent for the non-delayed cohort are for persons with the base characteristics. As with the loss probability analysis, the reenlistment probability indicated for each of the variables is for the characteristic defined by that variable substituted for its counterpart in the base.

The education, age, and mental group characteristics have probabilities ranging between 26 and 33 percent for the delayed cohort, and 12 through 17 percent for the non-delayed, indicating relatively little dependence on these characteristics. The reenlistment probabilities tend to decrease with higher educational attainment and higher mental group standing, presumably because higher quality enlistees expect to have better civilian job opportunities. Those with dependents exhibit appreciably larger reenlistment probabilities perhaps reflecting less willingness to accept the risks of a job search in the civilian sector. Similarly, the higher reenlistment probabilities for non-whites might result from their perception of better opportunities in the Navy.

Among the activity type variables, reenlistment probabilities range from 18 to 48 percent for the delayed cohort, and from 8 to 29 percent for the non-delayed; these variables thus account for a greater part of the variation than the background characteristics. Those assigned to submarines have the highest reenlistment rates, indicating a selection process that assigns highly motivated recruits to this activity type. On the other hand, enlistees in construction battalions exhibit the lowest reenlistment probabilities, presumably reflecting civilian job opportunities for such individuals. Note also the high reenlistment rates for the land-based air and "OTHER" activity types. Individuals in these activity types tend to be stationed on shore; as a result, the absence of the hardships associated with sea duty seems to increase reenlistment rates.

Overall, reenlistment probabilities for the school-delayed cohort are about twice as high as for the non-delayed. Within the two cohorts, the probabilities for different characteristics are related in similar ways; for example, each of the groups has a higher reenlistment probability associated with ED<11 relative to ED>12. This result arises because the two cohorts, pooled for the probit analysis, were differentiated only by the introduction of the school delay dummy variable, D; moreover, selectivity bias did not enter the calculation since it was found to play an unimportant role. In contrast, the loss probabilities shown in table 12, differentiate between delayed and non-delayed cohorts, and loss probabilities tend to increase for greater educational attainment in the delayed cohort while decreasing in the non-delayed. In that analysis, each of the cohorts was analyzed separately with no assumption of equality of coefficients for the two cohorts; in addition, selectivity bias was found to play an important role.

CONCLUSIONS

The previous sections provide evidence on the basis of which implications for Class A school admissions policies can be formulated. In what follows, we assume that the relationships established for the 1973 cohort of recruits will be valid for cohorts in the present and near future. Unless in the period 1973-1978 rather drastic changes have taken place in the unmeasured characteristics of enlistees and their living and working conditions in the Navy, this assumption is reasonable.

The main findings in this report that have implications for Class A school admissions policies are:

1. On the average, school delayed recruits are of lower quality (as measured by mental group and educational attainment) than the non-delayed.

2. A disproportionate fraction of recruits who are school delayed serve their first tour of duty on aircraft carriers.

3. Service time (defined as the elapsed time between completion of the first A school and the termination of service or the successful completion of four years) averages about 25 months for the school delayed and 37 months for the non-delayed.

4. The overall reenlistment rate (41 percent) for school-delayed recruits is about twice as high as for the non-delayed (21 percent).

5. Delayed enlistees of lower quality tend to have lower loss probabilities compared to their non-delayed counterparts. In particular, delayed individuals in mental group II with eleven or less years of education and mental group IV with twelve or more years of education have substantially lower loss probabilities.

6. For non-delayed recruits, the loss probability rises as their quality decreases. In particular, the non-delayed recruits with eleven or less years of education (about 15 percent of the cohort) have loss probabilities approximately 15 percentage points higher than the base group.

7. Reducing the time delay from about 18 months to 12 months increases the probabilities of loss by about 13 percent.

8. School delayed recruits whose first tour of duty was activity type "OTHER" (principally ashore and of relatively short duration) have appreciably higher loss probabilities than those whose first tour of duty was elsewhere.

REFERENCES

1. William Evanco, "Selectivity Bias in the Analysis of Military Manpower and Personnel Problems," Professional Paper 247, Unclassified, Center for Naval Analyses, Apr 1979

2. Jean W. Fletcher, "Estimating Four-Year Survival and Reenlistment Probabilities for Boiler Technicians and Machinist's Mates," Research Contribution 366, Unclassified, Center for Naval Analyses, Feb 1979

APPENDIX A

DESCRIPTION OF PROBIT ANALYSIS

APPENDIX A

DESCRIPTION OF PROBIT ANALYSIS

Suppose for each enlistee the Navy incurs costs C^* and accrues benefits B^*. The costs include base pay, training, various allowances, transportation, subsidies on purchases through Navy exchanges, health care, legal aid and counseling, as well as various indirect costs associated with personnel management and administration. In return, the Navy receives benefits that depend upon the enlistee's job performance. High levels of motivation, aptitude, education, and skill are generally associated with higher productivity. On the other hand, disciplinary and health problems not only reduce productivity, but can also increase the costs of maintaining an enlistee. Additional costs may be due to incarceration, court martial, or the need to recover an individual absent without leave.

If, for a given enlistee, the costs incurred, C^*, exceed the benefits B^*, then, assuming the Navy to be a rational decision- maker, the individual would be discharged. On the other hand, when the benefits exceed costs, the Navy would retain the individual. These two statements can be expressed mathematically. Define the quantity $L^* = C^*-B^*$. If costs exceed benefits, $L^* > 0$, and the enlistee should be lost; if benefits equal or exceed

cost, $L^* \leq 0$, and the individual should be retained.[1]

The actual measurement of costs and benefits for a given enlistee would be an extremely complex (and perhaps impossible) analytical task, but they can be related, although imperfectly, to the enlistee's attributes. Such measurable characteristics as age, educational level, intelligence, race, and numbers of dependents are thought to influence both job performance and costs. Other characteristics such as motivation, dependability, and initiative, may not be measurable, but are thought to be related to the measurable characteristics.

Let variables X_i, $i=1, \ldots, n$, represent the characteristics of an enlistee. Then, since costs and

---

[1] Several criticisms may be leveled at the above analysis. First, the benefits of a given individual may exhibit substantial externalities; the job performance of person A may influence the performance of person B. This situation might be especially strong in the Navy where teamwork is an important part of the working environment. Similarly, the costs must be viewed in global terms; for a disciplinary problem, the Navy must weigh the consequences of early discharge for an individual vs. prosecution on the rest of its personnel. Thus, for example, during hostilities, the Navy may choose to vigorously prosecute an individual absent without leave rather than resort to early discharge even though the costs of prosecution may far exceed the benefits of retaining the individual. Such a course of action may reduce the propensity of other personnel to be absent without leave so that the total benefits may exceed the costs of prosecution. While the formalism incorporating externalities could be developed, nothing substantive will be gained in the analysis by doing so.

Another criticism is that the benefits may not be measurable in pecuniary terms. The benefit, for example, might be measured as the contribution of the recruit to the mission of the Navy. If this were the case, then the above analysis can proceed on the basis of benefit-cost ratios whereby an individual is discharged if the ratio falls below a certain predetermined level and is retained if the level is exceeded.

Finally, as will be shortly demonstrated, the costs and benefits, <u>in practice</u>, are not required to be directly measurable. The only requirement for this analysis is that benefits and costs be measurable <u>in principle</u>.

benefits depend upon these characteristics, and the quality L^* depends upon costs and benefits, it is clear that L^* can be represented as a function of these characteristics. In particular, we assume that L^* can be expressed as a linear combination of the characteristics as follows:

$$L^* = \beta_o + \beta_1 X_1 + \beta_2 X_2 + \ldots + \beta_n X_n - u$$

Equivalently:

$$L^* = X\beta - u$$

where X is an (n+1)-dimensional row vector with $X_o = 1$ and β is an (n+1)-dimensional column vector.

The error term u results in part from the fact that measurable characteristics are an imperfect indicator of costs and benefits, and hence of L^*. This term is assumed to be normally distributed with zero mean, and incorporates unmeasured variables such as motivation, etc., in addition to errors in the measurement of L^*.

Given the characteristics associated with each individual in a cohort, this information can be used to estimate the values of the coefficients β_i, i=0, ..., n. If the value of L^* were known for each cohort member, then the β_i values could be obtained from linear regression analysis. We have no way to quantify L^*, but we will use retention data to infer whether L^* was greater than or less than zero. If an enlistee was discharged, we assume that costs exceeded benefits and L^* was greater than zero; on the other hand, we assumed that for any individual who was retained, L^* was at most zero.

For each cohort member, we define the observable variable L, whose values are related to L^* as follows:

$$L = 1 \text{ if and only if } L^* > 0 \text{ (a loss)}$$
$$L = 0 \text{ if and only if } L^* \leq 0 \text{ (a survivor)}$$

From equation (A1) we can infer that:

$$\begin{aligned} L = 1 &\text{ if and only if } u < X \\ L = 0 &\text{ if and only if } u \geq X \end{aligned} \qquad \text{(A2)}$$

The error term, u, shown in figure A-1 is normally distributed with mean zero. Without loss of generality, we can further assume that the standard deviation of u

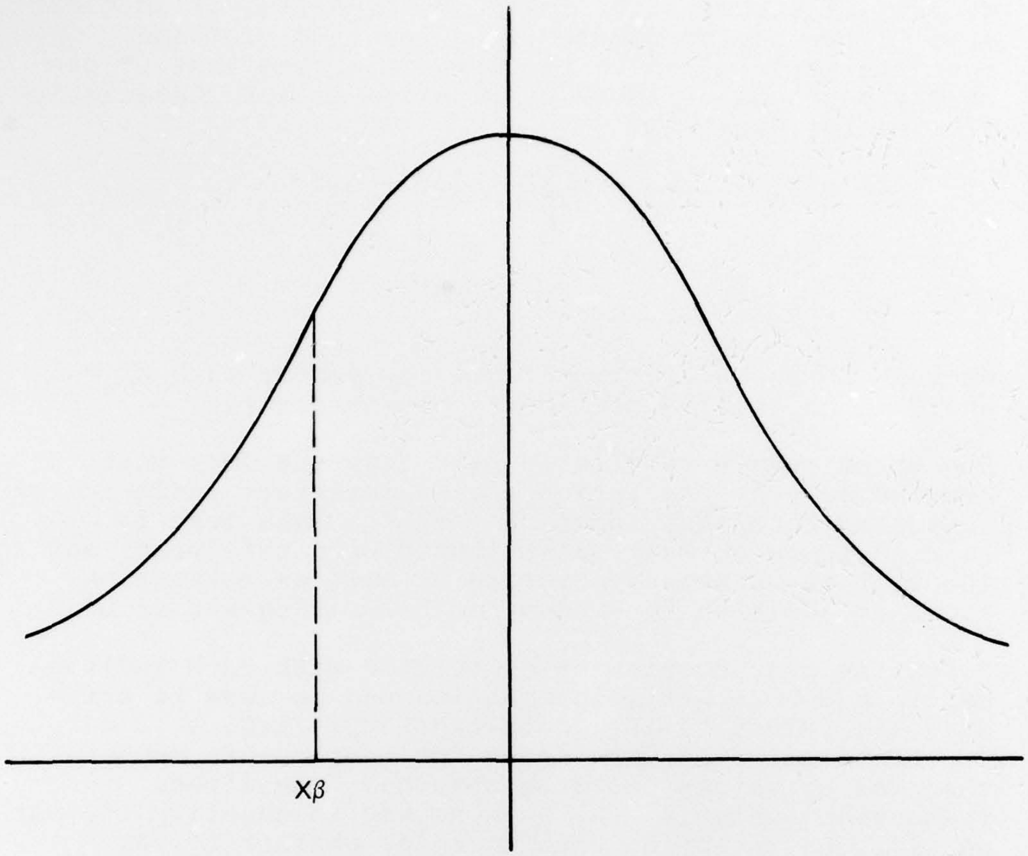


FIG. A1: NORMAL DISTRIBUTION WITH ZERO MEAN AND UNIT STANDARD DEVIATION

equals 1 since, otherwise, the inequalities in (A2) can be divided by σ to yield an error term with unit standard deviation.

In figure A-1, the probability of a loss (L=1), conditional upon characteristics X, is the area of the curve to the left of X, written mathematically as:

$$\text{Prob } (L=1 | X) = \int_{-\infty}^{X\beta} du\ f(u) = F(X\beta) \qquad \text{(A3)}$$

where f(u) is the normal distribution with zero mean and unit variance defined by:

$$f(u) = \exp(-u^2/2)/(2\pi)^{1/2} \qquad \text{(A4)}$$

and $F(X\beta)$ is the integral of the normal distribution of mean zero and unit variance between $-\infty$ and $X\beta$. Similarly, the probability of retention, given characteristics X, is:

$$\text{Prob } (L=0 | X) = \int_{X\beta}^{\infty} du\ f(u) = 1-F(X\beta) \qquad \text{(A5)}$$

The value of the β_i, $i=0, \ldots, n$, chosen are those which maximize the likelihood of the particular distribution of losses and retentions observed. The maximization process results in a system of equations non-linear in the β's. These can be solved by the Newton-Raphson technique, details of which can be found elsewhere [1].

The measured characteristics also include variables that depend upon selection and screening by the Navy or on choices made by the enlistee. These variables are functions of personal characteristics. For example, A school attendance depends upon Navy policy regarding school eligibility, which favors the higher mental groups.

Similarly, delayed entry to A school after recruit training may also be a function of personal characteristics. In a manner comparable to the analysis of loss probabilities, a variable D can be defined such that

$$\begin{aligned} D &= 1 \quad \text{for delayed training} \\ &= 0 \quad \text{for non-delayed training} \end{aligned}$$

This observable variable is based on an assumed underlying continuous variable D^* depending upon a recruit's characteristics such that

$$D = 1 \quad \text{if and only if} \quad D^* > 0$$
$$= 0 \quad \text{if and only if} \quad D^* \leq 0$$

where

$$D^* = X\delta - u_D$$

As before, X is a row vector of characteristics, δ is a column vector of coefficients and u_D is an error term incorporating missing variables and errors in the measurement of D^*. Thus, the probabilities of training delay and non-delay given characteristics X, are respectively:

$$\text{Prob}\ (D=1 | X) = F(X\delta) \tag{A6}$$
$$\text{Prob}\ (D=0 | X) = 1-F(X\delta) \tag{A7}$$

where $F(X\delta)$ is defined in (A3), and $f(u)$ is defined in (A4).

Characteristics such as A-school completion and delayed training, which are based on selectivity factors, may influence the probability of loss. On a priori grounds an individual who has completed A-school might be expected to exhibit a lower loss probability than others, all else being equal, since his working conditions are likely to be better than those of general detail. Similarly, within the cohort of individuals having completed A-school we might anticipate differential loss probabilities between delayed and non-delayed training, other factors being held constant. Since the individuals in the school delay group demonstrated perseverance by surviving general detail, and were screened on the basis of job performance, their loss probabilities may be lower than those of enlistees with non-delayed training, and similar personal characteristics.

To account for the difference in loss probabilities because of a selectivity factor such as training, we define the quantity L^* for the non-delayed and delayed training cohorts, respectively, as:

$$L^* = X\beta_0 - u_0 \quad \text{if } D = 0 \tag{A8}$$
$$L^* = X\beta_1 - u_1 \quad \text{if } D = 1 \tag{A9}$$

where X is a row vector of personal characteristics, as defined previously, and β_0 and β_1 are column vectors

of coefficients. The error terms u_0 and u_1 are assumed to be distributed normally with mean zero and unit standard deviations. If training delays affect loss probabilities, we would expect

$$\beta_0 \neq \beta_1 .$$

A probit analysis could be applied to equations (A8) and (A9) separately yielding results comparable to equations for the delayed and non-delayed training groups. However, such an approach might be inappropriate, because the factors involved in selectivity no longer guarantee that the distributions for u_0 and u_1 are normal. To pursue the analysis further define the joint theoretical distribution of u_0 , u_1, and u_D as $f(u_0, u_1, u_D)$ which is trivariate normal with variance-covariance matrix

$$\begin{pmatrix} 1 & \sigma_{01} & \sigma_{0D} \\ \sigma_{10} & 1 & \sigma_{1D} \\ \sigma_{D0} & \sigma_{D1} & 1 \end{pmatrix}$$

In addition, define the marginal distributions $f(u_0, u_D)$, which is bivariate normal with variance-covariance matrix

$$\begin{pmatrix} 1 & \sigma_{0D} \\ \sigma_{D0} & 1 \end{pmatrix}$$

and $f(u_1, u_D)$ also bivariate normal with variance-covariance matrix

$$\begin{pmatrix} 1 & \sigma_{1D} \\ \sigma_{D1} & 1 \end{pmatrix}$$

Then, the loss probabilities with and without training-delay given the characteristics X and the school-delay variable D, are, respectively:

$$\text{Prob}(L=1|\,X,D=1) = \frac{\int_{-\infty}^{X\beta_1} du_1 \int_{-\infty}^{X\gamma} du_D f(u_1, u_D)}{F(X\gamma)}$$

$$\text{Prob}(L=1|\ X,D=0) = \frac{\displaystyle\int_{-\infty}^{X\beta_0} du_0 \int_{X\gamma}^{\infty} du_D f(u_0,u_D)}{1-F(X\gamma)}$$

where the denominators are defined in (A6) and (A7). Formally, these equations can be rewritten as:

$$\text{Prob}(L=1|\ X,D=1) = \int_{-\infty}^{X\beta_1} du_1 g_1(u_1|\ X\gamma) \qquad \text{(A10)}$$

$$\text{Prob}(L=1|\ X,D=0) = \int_{-\infty}^{X\beta_0} du_0 g_0(u_0|\ X\gamma) \qquad \text{(A11)}$$

where

$$g_1(u_1|\ X\gamma) \int_{-\infty}^{X\gamma} du_D \frac{f(u_1,u_D)}{F(X\gamma)} \qquad \text{(A12)}$$

$$g_0(u_0,X\gamma) = \int_{X\gamma}^{\infty} du_D \frac{f(u_0,u_D)}{1-F(X\gamma)} \qquad \text{(A13)}$$

Equations (A12) and (A13) are not normal distributions in u_1 and u_0 respectively so that the usual univariate probit analysis cannot be applied to equations (A10) and (A11). However, as is shown in annex A, (A10) and (A11) can be approximated by:

$$\text{Prob } (L=1|\ X,D=1) = F(X\beta_1^* + \sigma_{1D}^* W_1) \qquad \text{(A10a)}$$

$$\text{Prob } (L=1|\ X,D=0) = F(X\beta_0^* - \sigma_{0D}^* W_0) \qquad \text{(A11a)}$$

where

$$\beta_i^* = \beta_i \,/(1 - \sigma_{iD}^2)^{1/2} \qquad \text{(A14)}$$

$$\sigma_{iD}^* = \sigma_{iD}/(1 - \sigma_{iD}^2)^{1/2} \qquad \text{(A15)}$$

and

$$W_0 = \frac{f(X\gamma)}{1-F(X\gamma)}$$

$$W_1 = \frac{f(X\gamma)}{F(X\gamma)}$$

Selectivity bias can then be corrected by means of the following two stage approach:

1. Using probit analysis, estimate the probability of school delay given by Prob(D=1 |X) in (A6) for the cohort of A school graduates as a function of background characteristics.

2. Next apply probit analysis to the delayed and non-delayed cohorts separately to estimate the loss probability based on the background characteristics, the service characteristics and variable W_1 for the school-delayed group and W_0 for the non-delayed. These expressions are given by (A10a) and (A11b).

ANNEX A-1

A BIVARIATE PROBIT MODEL

This annex presents the derivation of the bivariate probit model appropriate for analyzing problems involving a dichotomous endogenous variable that depends upon another dichotomous variable subject to selectivity. The study of loss probabilities uses this model.

Suppose continuous random variables Y^* and Z^* can be expressed in terms of exogenous variables such that:

$$Y_n^* = X_{1n}\,\beta_1 - u_{1n} \text{ if and only if } Z_n^* \geq 0 \qquad \text{(A-1-1a)}$$

$$= X_{2n}\,\beta_2 - u_{2n} \text{ if and only if } Z_n^* < 0 \qquad \text{(A-1-1b)}$$

and

$$Z_n^* = X_{3n}\gamma - u_{3n} \qquad \text{(A-1-2)}$$

where X_{in} (i=1,2,3) represents a 1 X K_i row vector of bounded variables and u_{1n}, u_{2n}, and u_{3n} are assumed to be normally distributed error terms for the nth observation, n=1,...N. β_1, β_2, and γ are K_ix1 (i=1, 2,3, respectively) column vectors of coefficients, which are estimated. We further assume that Y_n^* and Z_n^* are not directly observable but generate discrete endogenous variables Y_n and Z_n, respectively, defined by

$$Y_n = 1 \quad \text{iff } Y_n^* > 0 \qquad \text{(A-1-3a)}$$
$$= 0 \quad \text{otherwise}$$

$$Z_n = 1 \quad \text{iff } Z_n^* > 0 \qquad \text{(A-1-3b)}$$
$$= 0 \quad \text{otherwise}$$

where the thresholds are taken at zero without loss of generality. The error terms are assumed to satisfy

$$E(u_{in}) = 0$$

$$E(u_{in}^2) = 1$$

$$E(u_{in}u_{jn'}) = \sigma_{ij} \quad \text{if } n=n'$$
$$= 0 \quad \text{otherwise}$$

where i, j=1,2,3 and $i \neq j$. The variances are assumed equal to unity without loss of generality, since Y^*_n and Z^*_n can be defined to within arbitrary multiplicative constants.

Let $f(u_1,u_2,u_3)$ be the joint probability distribution of the error terms and $f(u_i,u_3)$, i=1,2, and $f(u_3)$ the marginal probability densities, defined by

$$f(u_1,u_3) = \int_{-\infty}^{\infty} du_2 \, f(u_1,u_2,u_3) \qquad \text{(A-1-4a)}$$

$$f(u_2,u_3) = \int_{-\infty}^{\infty} du_1 \, f(u_1,u_2,u_3) \qquad \text{(A-1-4b)}$$

$$f(u_3) = \int_{-\infty}^{\infty} du_1 \int_{-\infty}^{\infty} du_2 \, f(u_1,u_2,u_3) \qquad \text{(A-1-4c)}$$

Thus $f(u_i,u_3)$ is a bivariate normal distribution with covariance σ_{i3} and $f(u_3)$ is a univariate normal function defined by

$$f(t) = (1/2\pi)^{1/2} \exp(-t^2/2). \qquad \text{(A-1-5)}$$

From (3b) we see that for Z=1, $Z^* > 0$ or equivalently from (2), $u_3 < X_3\gamma$; thus since u_3 is normally distributed the probability for Z=1 is given by

$$\text{Prob}(Z=1) = F(X_3\gamma) \qquad \text{(A-1-6)}$$

where by definition

$$F(L) = \int_{-\infty}^{L} dt f(t) \qquad \text{(A-1-7)}$$

and f(t) is defined in (A-1-5). Similarly, because the sum of the probabilities for Z=0 and 1 must equal unity, the probability that Z=0 is

$$\text{Prob}(Z=0) = 1-F(X_3\gamma) = \int_{X_3\gamma}^{\infty} dt f(t) \qquad \text{(A-1-8)}$$

The probabilities for the occurrence of combinations of Y=0,1 and Z=0,1 are given by

$$\text{Prob}(Y=1,Z=1) = \int_{-\infty}^{X_1\beta_1} du_1 \int_{-\infty}^{X_3\gamma} du_3 f(u_3,u_3) \qquad \text{(A-1-9a)}$$

$$\text{Prob}(Y=0,Z=1) = 1-\text{Prob}(Y=1,Z=1) \qquad \text{(A-1-9b)}$$

$$\text{Prob}(Y=1,Z=0) = \int_{-\infty}^{X_2\beta_2} du_2 \int_{X_3\gamma}^{\infty} d u_3 f(u_2,u_3) \qquad \text{(A-1-9c)}$$

$$\text{Prob}(Y=0,Z=0) = 1-\text{Prob}(Y=1,Z=0) \qquad \text{(A-1-9d)}$$

The likelihood function, V, given particular values of Z_n, n=1,...,N is

$$V = \prod_{n=1}^{N} \text{Prob}(Y_n=1|\, Z_n=1)^{Y_n Z_n} \; \text{Prob}(Y_n=1|\, Z_n=0)^{Y_n(1-Z_n)}$$

(A-1-10)

$$\text{Prob}(Y_n=0|\, Z_n=1)^{(1-Y_n)Z_n} \; \text{Prob}(Y_n=0|\, Z_n=0)^{(1-Y_n)(1-Z_n)}$$

where the conditional probability, expressed in terms of equations (A-1-6), (A-1-8), and (A-1-9a)-(A-1-9d), is given by

$$\text{Prob}(Y_n|\, Z_n) = \text{Prob}(Y_n,Z_n)/\text{Prob}(Z_n) \qquad \text{(A-1-11)}$$

Thus the problem reduces to choosing values of the coefficients β_1, β_2 and γ for which V is a maximum. Since the likelihood function (A-1-10) is non-linear in these coefficients an approximation such as the

Newton-Raphson technique must be used to find a solution.

It is possible, however, to approximate (A-1-10), allowing it to be expressed as a univariate probit model for which computer algorithms are readily available. Consider the particular term Prob(Y_n=1| Z_n=1). Substituting (A-1-6) and (A-1-9a) into (A-1-11), and using the relationship

$$f(u_{1n},u_{3n}) = f(u_{3n})f(u_{1n}|\, u_{3n}) \qquad \text{(A-1-12)}$$

where, by recasting terms, we find that

$$f(u_{1n}|\, u_{3n}) = f(t)/(1-\sigma_{13}^{2})^{1/2} \qquad \text{(A-1-13)}$$

with f(t) defined in (A-1-5) and t defined by

$$u_{1n} = (1-\sigma_{13}^{2})^{1/2}\, t + \sigma_{13}u_{3n}, \qquad \text{(A-1-14)}$$

we get

$$\text{Prob}(Y_n=1|\, Z_n=1) = \int_{-\infty}^{X_{3n}\gamma} du_{3n} f(u_{3n}) \frac{F(X_{1n}\boldsymbol{\beta}_1^{*} - \sigma_{13}^{*}u_{3n})}{F(X_{3n}\gamma)} \qquad \text{(A-1-15)}$$

The functions F in the integrand are defined by equation (A-1-7). $F(X_{1n}\boldsymbol{\beta}_1^{*} - \sigma_{13}^{*}u_{3n})$ can be expanded in a Taylor series about u_{3n} = 0 and the moments of u_{3n} for the normal function $f(u_{3n})$ constrained between $-\infty$ and X_{3n} can be obtained; we can then write equation (15) as

$$\text{Prob}(Y_n=1 \mid Z_n=1) = F\left(X_{1n}\boldsymbol{\beta}_1^{*} + \sigma_{13}^{*} \frac{f(X_{3n}\gamma)}{F(X_{3n}\gamma)} + \sigma_{13}^{*} R(X_{1n}\boldsymbol{\beta}_1^{*}, X_{3n}\gamma)\right) \qquad \text{(A-1-15a)}$$

where

$$R(X_{1n}\boldsymbol{\beta}_1^{*}, X_{3n}\gamma) = -\frac{1}{F(X_{3n}\gamma)} \sum_{k=2}^{\infty} \frac{(\sigma_{13}^{*})^{k-1}}{k!} \qquad \text{(A-1-16)}$$

$$H_{k-1}(X_{1n}\beta_1^*)\mu_k(X_{3n}\gamma)$$

In the above expression, the H_k are Hermite polynomials defined iteratively by

$$H_{k+1}(z) = zH_k(z) - H_k'(z) \qquad \text{(A-1-17a)}$$

with

$$H_0(z) = 1 \qquad \text{(A-1-17b)}$$

and

$$\mu_k(X_{3n}\gamma) = \int_{-\infty}^{X_{3n}\gamma} dt\ t^k f(t) \qquad \text{(A-1-18)}$$

Keeping only the first order term of σ_{13}^* in (A-1-15a) yields:

$$\text{Prob}(Y_n=1|\ Z_n=1) = F\left(X_{1n}\beta_1^* + \sigma_{13}^* \ \frac{f(X_{3n}\gamma)}{F(Z_{3n}\gamma)}\right) \qquad \text{(A-1-15b)}$$

For $Z_n=0$, a comparable calculation yields

$$\text{Prob}(Y_n=1|\ Z_n=0) = F\left(X_{1n}\beta_0^* - \sigma_{23}^* \ \frac{f(X_{3n}\gamma)}{1-F(X_{3n}\gamma)}\right) \qquad \text{(A-1-15c)}$$

APPENDIX B

DATA BASE CREATION

APPENDIX B

DATA BASE CREATION

The data base used to predict retention and reenlistment rates is a longitudinal history of the Calendar Year 1973 cohort. Background information was extracted from the SCAT data tape created by AFEES for each non-prior service USN recruit entering the Navy in CY 1973. This information was merged with extracts of the June and December Enlisted Master Records (EMR) from 1973 to 1977. Loss data was incorporated from MARDAC and BUPERS loss tapes and schooling information was gathered from the NPRDC (Naval Personnel Research and Development Center) tapes. Finally, information from the Navy change tape for the first four months of 1978 was added to ensure identification of non-broken service reenlistments. The resulting data base is a complete longitudinal history for the CY 1973 cohort.

The sample used in the analysis includes all non-prior service USN male recruits who entered in CY 1973 and completed at least one Class A school. Men who attended, but did not complete A school are excluded from the sample. A school-delayed person is defined as one who returned from a non-training activity to attend and complete his first A school. The date received into the first regular tour of duty is compared with the starting date of the first A school completed; if the date received precedes the starting date, the man is classified as a school-delay. As an additional prerequisite, a school-delay is defined as such only if he does not have a special program code that guarantees schooling as a condition of enlistment. This eliminates individuals who attended class A school immediately following recruit training camp, failed to graduate, entered a non-training activity, and then returned to complete an A school.

The EMR extract and the NPRDC tape both provide school history data, although the NPRDC is more comprehensive. These data sources were compared to ensure consistency. Where obvious errors existed in one source, reasonable estimates were developed based on the alternative source. For example, many of the A school starting dates taken from the NPRDC tape were totally unreasonable. A technique was therefore developed to calculate average time in school for each rating. The appropriate average time was then subtracted from the A school completion date (taken from either the NPRDC or EMR extract tapes) to obtain an estimate of A school start-

ing date. This estimation technique was also employed when an individual had an EMR school history, but no matching NPRDC school history could be found. (EMRs do not record when an enlistee entered A school. When the completion dates of A school were discrepant but all other data matched and both dates appeared reasonable, the earlier of the two completion dates was chosen. This conservative approach biased the size of the school delay sample downward.

The starting date of the first regular tour of duty was determined by scanning the record of each man, starting with the onboard activity of the earliest EMR. If the activity type indicated a training activity, the scan was continued until the next EMR. This process continued until a non-training duty was obtained or the individual was lost (no more EMRs exist). The accounting category code was used to eliminate individuals whose only non-training activity was not a regular tour of duty; for example, some recruits were confined for medical or disciplinary reasons. When the first tour of duty had been located, the starting date of the tour was easily accessed from the EMR extract.

The definition of the activity types specified in table 2 can be found in reference [2].